Published in final edited form as:

Wiley Interdiscip Rev Syst Biol Med. 2010 ; 2(4): 438–459. doi:10.1002/wsbm.75.

In silico models of cancer

Lucas B. Edelman,

Institute for Genomic Biology, Department of Bioengineering, University of Illinois, Urbana-Champaign

James A. Eddy, and

Institute for Genomic Biology, Department of Bioengineering, University of Illinois, Urbana-Champaign

Nathan D. Price

Department of Chemical and Biomolecular Engineering, Institute for Genomic Biology, Center for Biophysics and Computational Biology, University of Illinois, Urbana-Champaign

Lucas B. Edelman: edelman2@illinois.edu; James A. Eddy: eddy2@illinois.edu; Nathan D. Price: ndprice@illinois.edu

Abstract

Cancer is a complex disease that involves multiple types of biological interactions across diverse physical, temporal, and biological scales. This complexity presents substantial challenges for the characterization of cancer biology, and motivates the study of cancer in the context of molecular, cellular, and physiological systems. Computational models of cancer are being developed to aid both biological discovery and clinical medicine. The development of these *in silico* models is facilitated by rapidly advancing experimental and analytical tools that generate information-rich, high-throughput biological data. Statistical models of cancer at the genomic, transcriptomic, and pathway levels have proven effective in developing diagnostic and prognostic molecular signatures, as well as in identifying perturbed pathways. Statistically-inferred network models can prove useful in settings where data overfitting can be avoided, and provide an important means for biological discovery. Mechanistically-based signaling and metabolic models that apply *a priori* knowledge of biochemical processes derived from experiments can also be reconstructed where data are available, and can provide insight and predictive ability regarding the dynamical behavior of these systems. At longer length scales, continuum and agent-based models of the tumor microenvironment and other tissue-level interactions enable modeling of cancer cell populations and tumor progression. Even though cancer has been among the most-studied human diseases using systems approaches, significant challenges remain before the enormous potential of *in silico* cancer biology can be fully realized.

Keywords

Systems medicine; Cancer; Personalized medicine; Systems biology; Computational biology

Motivations for In Silico Modeling in Cancer Research

Monumental advances in molecular and cellular biology – beginning in the latter half the 20th century and continuing today – have provided an increasingly detailed portrait of human biology from the molecular to physiological levels. These advances have centered on ‘reductionist’ experimental approaches aiming to annotate a vast array of biological components, from cells and tissues to genes and proteins. Collectively, these components represent a ‘parts list’ for biological systems (e.g., biochemical pathways, larger interaction networks). At scales beyond a handful of interacting components, however, simple analysis techniques can become limited in providing comprehensible insight into resulting

phenotypic behaviors. Systems biology is a rapidly growing discipline that employs an integrative approach to characterize biological systems, in which interactions among all components in a system are described mathematically to establish a computable model. These *in silico* models – which complement traditional *in vivo* animal models – can be simulated to quantitatively study the emergent behavior of a system of interacting components. Model development in the systems biology paradigm is enabled by the description of parts and interactions from reductionist biology, and also depends upon quantitative measurements. The advent of high-throughput experimental tools has allowed for the simultaneous measurement of thousands of biomolecules, paving the way for *in silico* model construction of increasingly large and diverse biological systems. Integrating heterogeneous dynamic data into quantitative predictive models holds great promise to significantly increase our ability to understand and rationally intervene in disease-perturbed biological systems. This promise – particularly with regards to personalized medicine and medical intervention – has motivated the development of new methods for systems analysis of human biology and disease.

Cancer is an intrinsically complex and heterogeneous disease, making it particularly amenable to systems biology approaches. Malignant tumors develop as a function of multiple biological interactions and events, both in the molecular domain among individual genes and proteins, and at the cellular and physiological levels between functionally diverse somatic cells and tissues [1] (Figure 1). At the molecular level, genetic lesions interact synergistically to evade tumor suppression pathways, with no single mutation typically sufficient to cause transformation [2–6]. Beyond genetic mutations, transformed cells can exhibit changes in expression of hundreds to thousands of genes and proteins [7–9]. Genetic modifications observed in cancer are often accompanied by changes at the epigenetic level [10–15]. The convolution of genetic effects and epigenetic modifications illustrates the complex, nonlinear relationship between molecular state and cellular cancer phenotype, emphasizing the need for heterogeneous data integration through *in silico* models. The diversity of cancer models mirrors the broad array of molecular and physiological events characteristic of the disease (Figure 2). The most course-grained approaches use statistical analysis of high-throughput expression data to identify molecular signatures of cancer phenotypes. Such signatures are indicative of aberrant function of genes or pathways, and can be used to predict the type, stage, or grade of biopsied tumor samples. More advanced methods aim to statistically infer the structure and/or quantitative relationships among biomolecules within interaction and regulatory networks of importance in cancer. Alternatively, stoichiometric or kinetic models of biochemical reaction networks – constructed in a bottom-up, annotation-based manner – can be used to simulate in mechanistic detail the behavior of metabolism or signal transduction in cancer.

The complexity of intracellular phenomena observed in cancer is mirrored by equally intricate interactions between cells and across somatic tissues. Among the most important biological systems mediating cancer development is the local tumor microenvironment – a complex, interacting system of cells and extracellular moieties. Contributory agents include the extracellular matrix [16], cooperating tumor and proximate ‘host’ cells [17–20], extracellular signaling factors [21–23], and the metabolic context of local tissue [24, 25]. Other important agents include the infiltrating leukocytes and cytokines of the immune system [26–28]. Human cancers also exhibit other major interactions with somatic tissues concomitant to malignant invasion, such as tumor-induced angiogenesis [29–31]. The potential response to chemotherapeutics, radiotherapy, and surgical procedures represent additional confounding factors in the cellular and physiological behavior of cancer cells [32–35]. The heterogeneous nature of the tumor microenvironment poses substantial modeling challenges, yet ongoing research has sought to characterize these cancer systems, including continuum and discrete models.

Despite significant experimental and analytic challenges arising from cancer's complexity, modeling has already successfully led to insights into cancer biology and treatment, as will be discussed herein. Some of the earliest models describing the molecular basis of cancer over half a century ago implicated the absolute number of genetic mutations as causative for malignancy [36]. Today, important efforts in sequencing the human genome [37, 38] and now individual cancers [39, 40] mean that malignant genetic transformations can be studied and modeled in the context of the entire genome. In this review, we describe key examples of recent *in silico* modeling efforts in cancer. These include (i) statistical models of cancer, such as molecular signatures of perturbed genes and molecular pathways, and statistically-inferred reaction networks; (ii) models that represent biochemical, metabolic, and signaling reaction networks important in oncogenesis, including constraint-based and dynamic approaches for the reconstruction of such networks; and (iii) continuum and agent-based models of the tumor microenvironment and tissue-level interactions. Finally, we examine the direction of ongoing research in cancer systems biology, and discuss opportunities for fundamental biological insights and their potential application to clinical practice.

Enabling Tools and Technology in Systems Cancer Modeling

With the initial sequencing of the Human Genome completed at the turn of the millennium, biological inquiry has been transformed by the ability to examine diverse cellular phenotypes such as cancer at the genome-scale. Experimental tools for measurement of the cancer genome, transcriptome, and proteome are being continuously improved, including microarray platforms [41–43], next-generation sequencing technologies [44, 45], and advanced mass spectrometry [46–48]. These tools for the high-throughput interrogation of cancer biology have developed in conjunction with analytical and computer-based tools for integration and annotation of the data generated. Together, these parallel advancements enable increasingly sophisticated models of cancer at the molecular and cellular levels.

Software and Database Resources for Cancer Systems Biology

With the rapid advancement of technologies for high-throughput biological measurement, the construction of tools for the distribution, integration and assessment of this raw information has become increasingly important [49]. Multiple online databases (Table 1) have been created for the storage and distribution of genome-scale data, including transcriptomics [50, 51], regulatory sequences [52, 53], and proteomics [54]. Importantly, these resources provide easily accessible high-throughput data by which cancer models can be constructed and validated. This access enables the global study of disease in a manner not previously possible and will continue to be transformative for systems biology and systems medicine. Models of cancer at different scales of size and complexity often employ *a priori* biological information as constraints to impose parsimony on model architecture. Other web-based services provide annotation of both individual genetic elements [55–59] and their collective associations within distinct biochemical pathways or functional modules [60–65]. Additionally, published mathematical models and required parameters (e.g., reaction kinetics) can be obtained for diverse biological systems [66–68]. Among the major challenges for both statistical and biochemical cancer modeling is knowledge of protein interactions [69]. Multiple public repositories of human and cancer protein interaction networks are currently available based on yeast-two-hybrid data and other experimental methods [70–73]. Additionally, standardized markup languages for biochemical and signaling networks [74–77], as well as software packages [78, 79] have been developed to enable rapid porting of diverse datasets, annotations, and models between different programs and formats. The Systems Biology Markup Language (SBML) in particular is a machine-readable format for representing models of biological processes that has been evolving over the past several years [76]. SBML is becoming more widely adopted, both in the reporting and distribution of models, and in the import and export capabilities of software toolboxes.

The application of advanced analytical techniques to the examination of high-throughput experimental data will continue to facilitate assembly of cancer models with increasing biological precision. However, several practical challenges remain to be addressed which impede the successful integration of data acquired from different experimental sources, or analyzed through different analytical methods. For example, a recent report found that curated protein interaction datasets integrating multiple literature sources can be highly error-prone [80]. The fusion of data across multiple studies, performed at different times or by different laboratory groups, poses a substantial challenge for systems-level analyses that rely so strongly upon large and information-rich datasets. The Minimum Information for Biological and Biomedical Investigations (MIBBI) project is an example of ongoing efforts focused on developing and sharing 'minimum information checklists' for experimental data. Such checklists place more stringent guidelines on the quality and content of published data [81]. Many journals and funding agencies already require authors reporting microarray data to comply with the Minimum information About a Microarray Experiment (MIAME) checklist [82]. Other minimal information guidelines which are gaining support in their respective communities include Minimum Information About a Proteomics Experiment (MIAPE) [83], Core Information for Metabolomics Reporting (CIMR), and Minimal Information About a RNAi Experiment (MIARE). The ability to estimate and compare the quality of different data integration schemes in the context of protein interactomics has also been the subject of recent investigation [84, 85]. As these challenges continue to be addressed going forward, they promise the development of ever more powerful and biologically accurate *in silico* representations of human cancer.

Statistically-derived Models of Cancer

Statistical models of cancer can be broadly divided into those that employ unbiased statistical inference, and those that also incorporate *a priori* constraints of specific biological interactions from data (Figure 3). Statistical models of cancer biology at the genetic, chromosomal, transcriptomic, and pathway levels provide insight about molecular etiology and consequences of malignant transformation despite incomplete knowledge of underlying biological interactions. These methods can help to elucidate key biomolecular events and pathways in oncogenic processes. They thus represent an important paradigm for both the fundamental characterization of cancer systems, and the discovery of molecular targets for diagnostic or therapeutic applications. Several studies have sought to infer the structure of small and large-scale biomolecular networks in human cells. These methods can describe previously known biological pathways, and can also be used to generate novel hypotheses about potential unknown interactions that can then be studied more thoroughly through experiments. However, multiple experimental and analytical challenges still remain in the statistical study of cellular and physiologic cancer systems, which must be overcome to yield increasingly useful insights into cancer biology.

Gene Expression Models for Cancer Diagnosis, Stratification, and Prognosis

The discrete mutational events found in the cancer genome and epigenome substantially modulate transcriptional profiles within cancer cells. Models based on these perturbed gene expression signatures can be applied for the diagnosis of disease subtypes, as well as for stratification of tumor grade or prediction of therapeutic response (Figure 3). These models apply deterministic or probabilistic rules to model the molecular context through which a cell or tissue progresses to malignancy. Different transcriptional classifiers have been developed for the discrimination of cancer type, subtype, and grade [89–91], including hierarchical clustering [92, 93], k-means clustering [94], support vector machines [95], artificial neural networks [96, 97], and classifiers based on the relative expression of gene pairs [98–101]. Transcriptomic signatures, which represent oncogenic processes as a function of gene expression levels, have also been applied to model relapse and overall

survival in diverse cancers, including non-small-cell lung cancer and pediatric leukemia [102–105]. Additionally, transcriptional classifiers have been used to predict tumor response to chemotherapeutic agents [106–108], which represents one of the most important clinical determinations in oncology practice.

These models often begin with genome-scale microarray data, and through either computational or combined computational and experimental analyses, derive prognostic classifiers comprised of a smaller number of highly relevant transcripts. The statistical analysis of molecular signatures has potential utility for informing small molecule, radiological, and surgical cancer treatment choices that cannot be filled by standard histopathological or clinical analyses. While these transcriptomic models have not entered widespread clinical use, they embody a promising direct application of cancer models to medicine. Furthermore, such models may be more widely adopted as nucleotide measurement technology itself becomes more common in non-research contexts. Transcriptional models of cancer also provide an accessible characterization of the molecular processes that mediate tumor initiation, progression, and sensitivity to medical interventions.

Pathway Enrichment Models and Biological Discovery

The genomics revolution has prompted substantial efforts towards comprehensively annotating biological functions of and associations among identified genes [58, 86]. Several curated databases have been established for grouping genes into modules of defined activity according to physical or functional association, such as Gene Ontology [61] and the KEGG database [88]. Multiple computational methods have been developed for the purpose of mapping genetic and transcriptional activity into these groups [109, 110]. These tools use diverse numerical methods to determine gene groupings whose activity, as represented by gene expression, is substantially modulated through malignant transformation. For example, Gene Set Enrichment Analysis [111] and related tools have been applied to identify pathway perturbations in human cancers based on transcriptomic data, including tumors of the breast and prostate [112–114]. An alternative approach combines human protein-protein interaction network data with gene expression profiles for the classification of breast and lung cancer samples [7, 115]. These related studies have considered the cumulative activity of discrete pathways, as defined by protein interaction data, to discriminate between phenotypes. This approach yielded results of similar predictive accuracy to the more common individual gene-based classifiers discussed previously, and demonstrates the importance of integrating various data sources for biological modeling. Similar pathway-based analysis methods also have been applied as prognostic indicators of clinical tumor progression – and the most important pathways identified through these computational methods often overlap significantly with those known experimentally to be oncogenic [116–118].

These frameworks have been applied to diverse cancer systems and serve as a robust source of biological discovery at the genomic scale. That said, mapping genetic activity into discrete, human-defined gene sets clearly does not accurately recapitulate the full biological complexity encoded by historical evolutionary constraints. Rather, it is a useful tool that leads to further, more detailed studies. Indeed, a large portion of fundamental eukaryotic cell biology remains poorly characterized. This includes complex events such as non-coding transcription and epigenetic phenomena, but also more basic information such as the specific components and mechanisms of various signaling cascades. As such, pathway annotations themselves are subject to incomplete information and biases in curation. Nonetheless, these techniques, while coarse-grained, enable the simultaneous consideration of many cellular pathways and processes using high throughput data.

Network-Based Statistical Inference

In silico cancer models based on statistical analysis of genetic aberrations, transcriptional activity, and pathway enrichment can provide important insights into the molecular phenomena that generate and propagate malignant cells. However, to characterize specific biomolecular mechanisms that drive oncogenesis, genetic and transcriptional activity must be considered in the context of cellular networks that ultimately drive cellular behavior [119–122]. In microbial cells, network inference tools have been developed and applied for the modeling of diverse biochemical, signaling, and gene expression networks [123–126]. Due to the much larger size of the human genome compared to microbes, and the substantially increased complexity of eukaryotic genetic regulation, inference of transcriptional regulatory networks in cancer presents increased practical and theoretical challenges. Additional complexity arises from the various data sources needed for studying human systems such as immortalized cell lines, primary cell lines, animal models, and tissue biopsies, which each provide imperfect recapitulation of *in vivo* cancer biology and varying degrees of experimental malleability. Nonetheless, several efforts to infer cancer networks – at both the level of individual pathways and at the genome scale – have revealed important information regarding the molecular basis of oncogenesis.

Network Inference at the Pathway Level

This scale of challenge to construct quantitatively predictive network models at the genome-scale in humans is not likely to be met in the near future [127], but more modest inference problems focusing on pathway-based subnetworks continue to prove useful. These studies examine different forms of biological data to infer causal relationships between genes at the transcriptional or signaling levels. For example, a multiparametric flow cytometry approach recapitulated a cell surface signaling pathway in the activation of human CD4+ leukocytes using a Bayesian network inference procedure [124]. Other studies have examined apoptotic signaling networks following experimental perturbation with pro-apoptotic TNF and other agents, using an innovative method in which downstream molecular perturbations are mapped into concordant ‘principal signaling components’ [128, 129]. These studies have shown that a small number of canonical signaling patterns can accurately describe even a complex cellular response such as apoptosis. Furthermore, by comparing signaling in different cell types with differing phenotypic responses, it was found that differential upstream signaling – at the level of kinase activation, for example – most strongly mediates cell-specific stimulus response, but that different signals are propagated through common intracellular networks.

A recent examination of receptor tyrosine kinases found that the complex and overlapping crosstalk involved in such signal transduction can be explained substantially by linear combinations for docking affinities for downstream proteins [130]. This study represents a systems-level characterization of an important cellular system that is too complex to be fully explicable using traditional ‘univariate’ biological modeling. Another recent study considered the cross-talk between phosphorylation pathways to predict the effects of inhibitor treatment on cancer cell migration phenotype [131]. This study characterized off-target and cross-talk effects of kinase inhibitor treatment to predict the effects of combinatoric treatments to forestall cell migration – an important malignant phenotype in cancer. Using different experimental and computational tools, these pathway-centric inference procedures have revealed important insights into diverse systems-level events in cancer that can have significant applications to pharmaceutical development and target selection.

These network inference studies incorporate *a priori* information in the design of the experiment itself – for example, in guiding the selection of what to measure (genes, proteins,

and/or phosphorylation events) and the planning of perturbation experiments to informatively test the system. This model-guided approach constitutes a reduction and focusing of network search space that increases the fidelity of network inference. These methods are not easily extendable to the inference of larger and more complex networks, as the larger number of variables introduces a much larger amount of possible interactions, with associated experimental noise and statistical uncertainty. As larger networks are examined, the task of ‘model identifiability’ becomes an increasingly refractory but critically important challenge.

Network Inference at the Genome Scale

The inference of biological networks at the genome scale involves several challenges for eukaryotic systems, and thus most network inference methods have been developed initially using unicellular organisms. One example is the ‘Inferelator’ computational suite that has been used to infer a genome-scale transcription regulatory network in the unicellular archaeon *Halobacterium* [132, 133]. This study constructed a network model of transcription able to determine the level of expression of each gene given the expression state of transcriptional regulators. This model was able to predict quantitatively the response of the organism to novel stimuli that were not used to train the model itself. This type of quantitative inferred genome-scale transcriptional regulatory network in humans remains a difficult and ongoing challenge.

Nonetheless, efforts have been made to craft network-based statistical models of human cells, including in breast cancer and lymphoma, in which the architecture of regulatory networks for a portion of the human genome is characterized [134–137]. These models have applied Bayesian network inference methods [138] to discriminate physical or functional interactions between several thousand genes. The related probabilistic Boolean network formalism [139] has been used to construct regulatory networks in the context of several cancer systems, including glioma [140]. One study of melanoma adapted Boolean network analysis to include the use of ‘seed’ genes, in which groups of genes believed to interact were used as a subnetwork starting point which was then iteratively grown to incorporate other related genes [141]. Many network inference methods appear robust when validated on synthetic data, but exhibit substantially reduced performance when applied to experimental gene expression datasets [142–144]. These results speak to the substantial complexity of gene regulation networks – even in the context of unicellular organisms – and to the requirement for large datasets in network inference tasks [145].

Statistically inferred network models can be used to study the topology of complex cellular systems, and elucidate important genetic interactions and control points. Such statistical networks map the status of regulatory agents into qualitative interactions with cognate targets and discrete network states. Going forward, the acquisition of larger high-throughput datasets encoding different facets of regulatory interactions, combined with innovative methods for their integration and analysis, will enable the construction of increasingly precise numerical network models. The important ability to reliably model the behavior of cancer cells, for example upon response to treatment or genetic evolution, would hold significant implications for the development and administration of cancer treatments.

Biochemical Reaction Networks

In contrast to statistically-inferred networks, biochemical reaction networks are constructed to represent explicitly the mechanistic relationships between genes, proteins, and the chemical interconversion of metabolites within a biological system (Figure 4). In these models, network links are based on pre-established biomolecular interactions rather than statistical associations; significant experimental characterization is thus needed to

reconstruct biochemical reaction networks in human cells. These biochemical reaction networks require, at a minimum, knowledge of the stoichiometry of the participating reactions. Additional information such as thermodynamics, enzyme capacity constraints, time-series concentration profiles, and kinetic rate constants can be incorporated to compose more detailed dynamic models.

Stoichiometric Models of Biochemical Reaction Networks

The most basic mathematical representation of a biochemical reaction network is a stoichiometric model. Stoichiometric models describe the interconversion of biomolecules purely in terms of the number of reactants and products participating in each reaction. The generation of stoichiometric and analysis of their properties is a well-established process [146–148], and genome-scale models of metabolism have been completed for a diverse range of organisms, including the prokaryote *E. coli* [149], archaea [150], and eukaryotes such as *S. cerevisiae* [151] and the protozoan *L. major* [152]. Most important for cancer research is the recent reconstruction of known human metabolism at the genome-scale [153, 154]. Methods have also been developed for reconstructing signaling networks [155–157], transcriptional and translational networks [146], and regulatory networks [158, 159]; these models are fundamentally analogous to reconstructed metabolic networks (Figure 5). The reconstruction of a biochemical reaction network results in a database of stoichiometric equations that can be represented mathematically to form the foundation of a genome-scale, computable model. Computational tools for constraint-based analysis are then used to interrogate the properties of the reconstructed network *in silico*, and to facilitate model-driven validation and refinement [160]. Physico-chemical and environmental constraints under which the network operates are applied in the form of *balances*, including mass, energy, and charge, and *bounds*, such as flux capacities and thermodynamic constraints. The statement of constraints defines a solution space comprising all non-excluded network states, thereby describing possible functions or allowable phenotypes. These methods are now being adapted for modeling human systems in greater detail.

Constraint-based analysis of biochemical reaction networks has been applied to a number of human systems. Using the reconstruction of the human mitochondrial metabolic network [161], linear programming and random sampling were applied to identify candidate steady states of the network under normal, diabetic, ischemic, and dietetic conditions [162]. In a related study, Monte Carlo sampling of flux spaces was used to study the effects of enzymopathies on the human erythrocyte metabolic network [163]. A stoichiometric formalism was also used to reconstruct reactions comprising the JAK-STAT signaling network in the human B-cell [164]. In this study, a topological constraint-based approach was employed to evaluate network crosstalk, redundancy in signaling inputs and outputs, and to delineate correlated reaction sets or systemic modules. The recent completion of a global reconstruction of the human metabolic network [153] represents a significant milestone in human systems biology – and clears a path forward for reconstructing metabolic models for all of the 200 cell types in the human body and their modified forms in various cancers. This reconstruction comprises 1496 genes and 3748 reactions divided into 88 metabolic pathways. In addition to the typical network capabilities determined by constraint-based modeling, the initial genome-scale reconstruction has enabled analysis of relationships between network topology and human metabolic diseases [165]. An independently reconstructed human metabolic network [154] – comprising 2322 and 2823 reactions among 70 different pathways – was also used to demonstrate the potential of systems modeling in human metabolism to aid in drug discovery [166]. An important effort going forward will be to resolve the discrepancies between different versions of the human reconstructions, establishing a consensus global reconstruction of human metabolism.

Modeling Metabolism in Human Cancer

The global human metabolic reconstruction provides a basis for the known set of metabolic reactions catalyzed by human proteins. However, the utility of these models for cancer research going forward depends upon overcoming several challenges. First, further refinement of the global human metabolic map is essential to increase its accuracy. Second, each of the approximately 200 cell types in the human body exhibits only a portion of the full metabolic capability contained in the genome [167] (Table 2). The high percentage of undetermined activities for metabolic enzymes in human tissues clearly shows how much more we have to learn about even this very well studied cellular process. Effectively representing which portions of the global human metabolic network are active in any given cell type, and at what level, is thus of critical importance. Cancers in particular are known to exhibit diverse metabolic phenotypes compared to their progenitor cells, typically with an increased rate of overall metabolic activity to support their increased growth and the highest metabolic activity observed in the most aggressive malignancies [168, 169]. Multiple other hallmarks of cancer including angiogenesis, metastasis, evasion of apoptosis, and avoidance of immune detection have been previously linked to human tumor metabolism [170, 171]. Metabolic targets have also been used in cancer chemotherapy [172–174]. For these reasons, metabolic networks in human cancer have the potential to be a rich focus area for systems modeling going forward.

Kinetic Models of Biochemical Reaction Networks

Detailed dynamic models are needed to accurately quantify many important molecular processes, such as feedback and feedforward regulation, competitive inhibition, post-translational modification, and transcriptional regulation. At scales much smaller than the genome, biochemical reaction networks have been used as the basis for forming dynamic differential equation models with the addition of kinetic rate constants [155, 175, 176]. Many of the fundamental considerations and challenges associated with constructing kinetic models of cell signaling pathways have been previously reviewed in detail [177, 178]. Briefly, these models are most commonly formulated as a set of ordinary differential equations (ODEs). In these ODE-based pathway models, the rates of production and consumption of individual biomolecular species is described in terms of mass action kinetics (elementary reactions with forward and reverse rate constants). As the full kinetic detail for reactions in biological pathways is often unknown, simplifying assumptions such as the Michaelis-Menten approximation or time-scale separation can be applied. Kinetic models are advantageous in that they do not employ a steady-state assumption as constraint-based models typically do, and thus can simulate detailed dynamic behavior. Dynamic models are also able to simultaneously account for both the concentrations of compounds and flux through reactions. Capturing dynamics however, requires more parameters, and thus these models are more data intensive to create, as well as more prone to overfitting when parameters must be estimated. Nevertheless, smaller scale dynamic models have been used successfully to study in mechanistic detail key pathways related to human cancer. Most of the pathways and molecular components listed below do not function independently to promote oncogenesis, but rather are often connected through signaling crosstalk, feedback mechanisms, and other forms of up- or downstream regulation. These examples are not an exhaustive list, but serve to demonstrate the benefits of detailed kinetic modeling of biological pathways.

Aberrant activity of the transcription factor NF- κ B has been linked to oncogenesis, tumor progression, and resistance to chemotherapy. A computational ODE model was employed to determine the role of inhibitor of NF- κ B kinase (I κ B) isoforms in the temporal control of NF- κ B [155]. Analysis of the model revealed that I κ B α is responsible for strong negative feedback and fast turn-off of the NF- κ B response to IKK stimulation, while I κ B β and I κ B ϵ

reduce oscillations in the signaling module and stabilize NF- κ B response during longer stimulation. The same model was expanded and used to study dynamics of I κ B–NF- κ B signaling module when cells were stimulated by lipopolysaccharide (LPS) via Toll-like receptor 4 (TLR4) [175, 179]. Modeling of two TLR4-dependent signaling pathways suggested that one pathway required the expression tumor necrosis factor- α (TNF α) to activate NF- κ B.

In another study, the negative feedback loop between the tumor suppressor p53 and the Mdm2 oncogene was examined experimentally and computationally [180]. Different mathematical models of the system were analyzed to draw connections between predicted and experimentally observed behavior. The models indicated that low-frequency noise in protein production rates was an important source of variability in oscillations of the p53–Mdm2 system. Another model was built for the single-cell response of p53 to radiation-induced DNA damage [181]. In agreement with experiments, simulations showed that by assuming stochasticity in the initial number of double-strand break sites and the DNA repair process, p53 and Mdm2 exhibit coordinated oscillatory dynamics with radiation stimulation.

Numerous components of the phosphatidylinositol-3-kinase (PI3K)/AKT pathway are targeted by malignant perturbations such as amplification, mutation, and translocation in cancer patients. Activation of the PI3K pathway results in aberrant cell growth and survival, metastatic competence, and therapy resistance. An excellent review of the challenges, opportunities, and progress in modeling and exploiting the PI3K pathway can be found in Hennessy et al., 2005 [182]. For example, inhibitors of the mTOR pathway have been demonstrated to directly target PI3K signaling. However, due to the intricate role of mTOR in regulation of PI3K/AKT signaling, detailed understanding of the pathway derived from computational modeling is needed to determine the consequences of inhibition of particular components.

A number of models have been used to examine mechanisms that govern mitogen-activated protein kinase (MAPK) pathway dynamics. MAPK pathways are involved in the activation of many cancer-promoting genes and respond to a number of different cancer-associated stimuli, such as epidermal growth factor (EGF). A computational model of general MAPK signaling pathways was used to study the role of negative feedback mechanisms in generating complete signal adaptation (i.e., desensitization to stimulus) [183]. Another model was used to simulate EGF-induced MAPK activation [184]. Analysis of model simulations provided quantitative evidence that feedback inhibition of the MAPK cascade is the most important factor in determining the duration of cascade activation. It was also found that differences in EGF and NGF-stimulated signaling could be accounted for by differential feedback regulation of the MAPK cascade. Similarly, a model was constructed to examine the signal-response relationships between EGF binding and the activation of downstream proteins in the cascade [185]. Model predictions showed that EGF-induced responses are remarkably stable over a 100-fold range of ligand concentration.

Another model used dynamic simulations of gene expression and signaling networks to study behavior in human cancer cells [186]. While dynamic interaction networks of cancer systems are challenging to construct and simulate, such models are essential to capture the molecular complexity of oncogenesis. Taken together, these studies provide powerful examples of how building a model that correctly explains data and makes verifiable predictions can lead to deepened insight into biological mechanism and function.

Microenvironment and Tissue-Level Models

In silico models of cancer can be built not only for intracellular networks, but also at larger length scales. Alternative computational methods must be applied to consider the interface

between cancers and the tissue contexts in which they reside. These settings exhibit complex interactions with multiple factors of different function and scale, including extracellular biomolecules, a spatially intricate and dynamic vasculature, and the immune system [187]. Models of cancer at the tissue level that account for these functionally divergent parameters can be broadly divided into ‘continuum’ models, and discrete or ‘agent-based’ models [188–190]. The latter are often applied when the number of individual interacting units, such as cancer cells, is constrained to remain small; the former is more practical at population scales where agent-based modeling can be computationally prohibitive. Both methods can integrate information about the biological context in which cancers develop, and thus represent a multi-scale consideration of oncogenesis as it occurs within somatic tissues [191].

Continuum Models

Extracellular parameters can be represented as continuously distributed variables to mathematically model cell-cell or cell-environment interactions in the context of cancers and the tumor microenvironment. Systems of partial differential equations have been used to simulate the magnitude of interaction between these factors, including the effects of hypoxia on cell cycle progression [192], the impact of mechanical forces on tumor invasiveness [193] and extracellular matrix interactions [194]. Recent studies have examined cell population dynamics within colonic crypts in colorectal cancer [195, 196]. These models consider interactions between stem cells, differentiating cells, and differentiated cell populations to quantitatively predict tissue-level invasion and the growth of tumor mass. Other models have represented solid tumors as a multiphase system of both bound and ‘mobile’ forms [197, 198]. Such ‘mixture’ models consider differential growth and apoptosis rates, as well as mass transfer and regulatory interactions between phases. Alternative models have considered nonlinear and combinatoric effects of multiple factors, including nutrient availability and mechanical parameters [199, 200], and the effects of mutation rate on invasion and metastasis [201]. These numerical systems embody a robust method to incorporate the effects of somatic biological phenomena into computational representations of cancer. Continuum-based models are thus a powerful tool to simulate and characterize interactions between intracellular and extracellular factors in oncogenic processes.

Cellular Automata and Agent-Based Models

Multivariate continuum models are able to represent the effects of several physiological or biochemical events on cancer development. However, *in situ*, these factors are highly heterogeneous, and interact discontinuously with tumor cells [202, 203]. Cellular automata models represent cancer cells as discrete entities of defined location and scale, interacting with one another and external factors in discrete time intervals according to predefined rules [204]. Agent-based models expand the cellular automata paradigm to include entities of divergent functionalities interacting together in a single spatial representation, including different cell types, genetic elements, and environmental factors [205]. With sensitivity to starting conditions, and the ability to incorporate probabilistic interactions at each time step, these models can exhibit similar stochastic behaviors to those observed in oncogenesis *in vivo* [206–208]. Phenomena that have been modeled using agent-based models include three-dimensional tumor cell patterning [209], immune system surveillance [210], angiogenesis [211–213], and the kinetics of cell motility [214]. Another recent model integrated diverse parameters such as extracellular signals, blood flow, and tissue degradation to simulate the spatiotemporal formation of tumor vasculature [215].

Increasingly, ‘hybrid’ models have been created which incorporate both continuum and agent-based variables in a modular approach. For example, a recent study considered continuous extracellular biomolecule distributions and discrete cell locations to simulate the

interaction between intracellular decision-making processes and malignant growth [216]. Another recent model incorporated a continuous model of a receptor signaling pathway, an intracellular transcriptional regulatory network, cell-cycle kinetics, and three-dimensional cell migration in an integrated, agent-based simulation of solid brain tumor development [217]. The interaction between cellular and microenvironment states have also been considered in a multiscale model that predicts tumor morphology and phenotypic evolution in response to such extracellular pressures [218]. These and other techniques which incorporate multiple, nested scales of interacting biology embody promising paradigms to understand cancer as a cascade of information across levels of size and complexity. This ability to interrogate cancer across multiple biological agents and compartments presents a unique framework to elucidate oncogenic processes, and to evaluate potential therapeutic interventions through digital simulation prior to experimental deployment.

Conclusion

The ability to model the initiation and progression of human cancer at the level of molecules, cells, and tissues is important to improve cancer diagnosis and treatment. We have discussed the biological motivations for *in silico* models of cancer with regards to the molecular complexity observed in the disease, which is best understood in the context of systems-level interactions. Similarly, tumors interface with an intricate environment of several cell types and extracellular factors, an additional layer of complexity which will require systems-based approaches to address reliably. Powerful new technologies enable the high-throughput biological measurements critical to understanding these processes, including nucleotide sequencing, proteomics, and protein interaction assays. Additionally, the standardization and distribution of this data through common file formats and internet databases empowers the research community to build efficiently upon previous studies and to cumulatively expand our knowledge of the disease. Predictive models serve as compact and rigorous representations of large-scale hypotheses for the functioning of disease-perturbed networks, and promise to advance rational therapy design in the coming future of systems medicine.

Diverse computational methods have been developed to analyze different elements of cancer biology, and to elucidate multicomponent, systems-level contributors to cancer development. Statistical models consider cancer cell biology at the pathway or genome scale to elucidate genetic or transcriptional signatures associated with oncogenesis. The ability to identify molecular signatures that can inform diagnosis and treatment selection is already beginning to take effect in the clinic, for example with molecularly targeted therapies such as imatinib (Gleevec) for selected subpopulations of cancer [219]. At a finer resolution, quantitative pathway-level models offer explicit representation of intracellular signaling and metabolic events in cancer, and important scientific successes have been observed with these mechanistic pathway models. Integration of many molecular pathways for quantitative genome-scale modeling remains an ongoing challenge, with statistical inference of cellular networks currently possible only in a very limited context in human cancers. At the physiological level, continuum and agent-based models efficiently describe the interaction between cancer cells and the surrounding tissue environment. Cumulatively, these models are able to integrate many different sources of experimental data to simulate and observe diverse aspects of cancer biology. Translation of these models to the clinic, while currently a slow process, suggests a substantial promise of systems medicine that is just beginning to be explored.

The future prospects for *in silico* models of cancer are substantial, and this field is clearly still in its infancy. With rapidly advancing technologies for biologic interrogation at the genetic, transcriptional, proteomic, and cellular levels, the body of data upon which models

can be trained is expanding rapidly. Coupled with the persistent exponential rise of computing power, the ‘raw materials’ of systems biology will continue to improve at an exhilarating rate. With the broad availability of annotations for cellular pathways and biomolecular interactions, specific constraints can now be imposed across the genome on model architectures to enhance dimensional parsimony and statistical resolution. Additionally, a broad array of tools for numerical simulation of both individual biological modules and genome-scale phenomena has been developed. Going forward, the fusion of experimental, informational, and analytical modalities will provide a powerful analytical framework for *in silico* modeling of human cancer.

In silico cancer modeling presents significant opportunities to investigate oncogenesis across biological scales and systems. We are optimistic that these powerful methods will help to accelerate the development of diagnostic and therapeutic technologies for clinical medicine. Considerable improvement in the resolution, scale, and predictive power of these models must first be achieved, regarding which substantial challenges remain. Nonetheless, it may ultimately be possible to simulate oncogenesis and malignant invasion accurately from the scale of genetics to physiology. With this technology, distinguishing signatures of cancers could be discovered automatically for purposes of early diagnosis, prognosis, and treatment planning. Additionally, and perhaps most promisingly, with reliable digital representations of cancer, the effects of therapeutic interventions at both the molecular and surgical scales could be predicted *in silico* without exposing patients to risk. This innovation would greatly accelerate the development of safe and targeted anticancer therapeutics, and offer hope of medical treatments for diseases that remain refractory to current clinical technologies. Ultimately, while confronting substantial experimental and analytical challenges, *in silico* models of cancer are advancing, and promise to strongly enhance both the fundamental understanding of cancer and its treatment in the clinic.

References

1. Hanahan D, Weinberg RA. The Hallmarks of Cancer. *Cell*. 2000; 100(1):57–70. [PubMed: 10647931]
2. Land H, Parada LF, Weinberg RA. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature*. 1983; 304(5927):596–602. [PubMed: 6308472]
3. Lloyd AC, Obermuller F, Staddon S, Barth CF, McMahon M, Land H. Cooperating oncogenes converge to regulate cyclin/cdk complexes. *Genes Dev*. 1997; 11(5):663–677. [PubMed: 9119230]
4. Fanidi A, Harrington EA, Evan GI. Cooperative interaction between c-myc and bcl-2 proto-oncogenes. *Nature*. 1992; 359(6395):554–556. [PubMed: 1406976]
5. Lowe SW, Cepero E, Evan G. Intrinsic tumour suppression. *Nature*. 2004; 432(7015):307–15. [PubMed: 15549092]
6. McMurray HR, Sampson ER, Compitello G, Kinsey C, Newman L, Smith B, Chen SR, Klebanov L, Salzman P, Yakovlev A. Synergistic response to oncogenic mutations defines gene class critical to cancer phenotype. *Nature*. 2008; 453(7198):1112. [PubMed: 18500333]
7. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol*. 2007; 3(140)
8. Liu ET, Lemberger T. Higher order structure in the cancer transcriptome and systems medicine. *Mol Syst Biol*. 2007; 3(94)
9. Auffray C. Protein subnetwork markers improve prediction of cancer outcome. *Mol Syst Biol*. 2007; 3(141)
10. Neely KE, Workman JL. The complexity of chromatin remodeling and its links to cancer. *Biochim Biophys Acta*. 2002; 1603(1):19–29. [PubMed: 12242108]
11. Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, Kurdistani SK. Global histone modification patterns predict risk of prostate cancer recurrence. *Nature*. 2005; 435(7046):1262. [PubMed: 15988529]

12. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *DNA*. 2007; 1(E2):E1.
13. Jones PA. DNA methylation and cancer. *Oncogene*. 2002; 21:5358–5360. [PubMed: 12154398]
14. Esteller M, Fraga MF, Paz MF, Campo E, Colomer D, Novo FJ, Calasanz MJ, Galm O, Guo M, Benitez J. Cancer epigenetics and methylation. *Science*. 2002; 297(5588):1807–8. [PubMed: 12229925]
15. Laird PW. Cancer epigenetics. *Hum Mol Genet*. 2005; 14(90001):65–76.
16. Tang Y, Nakada MT, Kesavan P, McCabe F, Millar H, Rafferty P, Bugelski P, Yan L. Extracellular Matrix Metalloproteinase Inducer Stimulates Tumor Angiogenesis by Elevating Vascular Endothelial Cell Growth Factor and Matrix Metalloproteinases. *Cancer Res*. 2005; 65(8): 3193–3199. [PubMed: 15833850]
17. Condeelis J, Pollard JW. Macrophages: Obligate Partners for Tumor Cell Migration Invasion, and Metastasis. *Cell*. 2006; 124(2):263–266. [PubMed: 16439202]
18. Lewis CE, Pollard JW. Distinct Role of Macrophages in Different Tumor Microenvironments. *Cancer Res*. 2006; 66(2):605–612. [PubMed: 16423985]
19. Kalluri R, Zeisberg M. Fibroblasts in cancer. *Nat Rev Cancer*. 2006; 6(5):392. [PubMed: 16572188]
20. Bhowmick NA, Neilson EG, Moses HL. Stromal fibroblasts in cancer initiation and progression. *Nature*. 2004; 432(7015):332–337. [PubMed: 15549095]
21. Sheu BC, Chang WC, Cheng CY, Lin HH, Chang DY, Huang SC. Cytokine regulation networks in the cancer microenvironment. *Front Biosci*. 2008; 13:6255–68. [PubMed: 18508658]
22. De Luca A, Carotenuto A, Rachiglio A, Gallo M, Maiello MR, Aldinucci D, Pinto A, Normanno N. The role of the EGFR signaling in tumor microenvironment. *J Cell Physiol*. 2008; 214(3):559–67. [PubMed: 17894407]
23. Allinen M, Beroukhi R, Cai L, Brennan C, Lahti-Domenici J, Huang H, Porter D, Hu M, Chin L, Richardson A. Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell*. 2004; 6(1):17–32. [PubMed: 15261139]
24. Smallbone K, Gavanagh DJ, Gatenby RA, Maini PK. The role of acidity in solid tumour growth and invasion. *J Theor Biol*. 2005; 235(4):476–484. [PubMed: 15935166]
25. Smallbone K, Gatenby RA, Gillies RJ, Maini PK, Gavanagh DJ. Metabolic changes during carcinogenesis: Potential impact on invasiveness. *J Theor Biol*. 2007; (2444):703–713. [PubMed: 17055536]
26. Gajewski TF, Meng Y, Harlin H. Immune Suppression in the Tumor Microenvironment. *J Immunother*. 2006; 29(3):233. [PubMed: 16699366]
27. Gajewski TF, Meng Y, Blank C, Brown I, Kacha A, Kline J, Harlin H. Immune resistance orchestrated by the tumor microenvironment. *Immunol Rev*. 2006; 213(1):131–145. [PubMed: 16972901]
28. Croci DO, Zacarias Fluck MF, Rico MJ, Matar P, Rabinovich GA, Scharovsky OG. Dynamic cross-talk between tumor and immune cells in orchestrating the immunosuppressive network at the tumor microenvironment. *Cancer Immunol Immunother*. 2007; 56(11):1687–1700. [PubMed: 17571260]
29. Laakkonen P, Waltari M, Holopainen T, Takahashi T, Pytowski B, Steiner P, Hicklin D, Persaud K, Tonra JR, Witte L. Vascular Endothelial Growth Factor Receptor 3 Is Involved in Tumor Angiogenesis and Growth. *Cancer Res*. 2007; 67(2):593. [PubMed: 17234768]
30. Su JL, Yen CJ, Chen PS, Chuang SE, Hong CC, Kuo IH, Chen HY, Hung MC, Kuo ML. The role of the VEGF-C/VEGFR-3 axis in cancer progression. *Br J Cancer*. 2007; 96:541–545. [PubMed: 17164762]
31. Hirakawa S, Brown LF, Kodama S, Paavonen K, Alitalo K, Detmar M. VEGF-C-induced lymphangiogenesis in sentinel lymph nodes promotes tumor metastasis to distant sites. *Blood*. 2007; 109(3):1010. [PubMed: 17032920]
32. Taranawski R, Skladowski K, Swierniaka A, Wygoda A, Mucha A. Repopulation of tumour cells during radiotherapy is doubled during treatment gaps. *Computational and Mathematical Methods in Medicine*. 2000; 2(4):297–305.

33. Kohandel M, Sivaloganathan S, Oza A. Mathematical modeling of ovarian cancer treatments: Sequencing of surgery and chemotherapy. *J Theor Biol.* 2006; 242(1):62–68. [PubMed: 16580022]
34. Verschraegen C, Vinh-Hung V, Cserni G, Gordon R, Royce ME, Vlastos G, Tai P, Storme G. Modeling the Effect of Tumor Size in Early Breast Cancer. *Ann Surg.* 2005; 241(2):309. [PubMed: 15650642]
35. Ayati BP, Webb GF, Anderson ARA. Computational Methods and Results for Structured Multiscale Models of Tumor Invasion. *Multiscale Modeling & Simulation.* 2006; 5:1.
36. Nordling CO. A new theory on cancer-inducing mechanism. *Br J Cancer.* 1953; 7(1):68–72. [PubMed: 13051507]
37. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W. Initial sequencing and analysis of the human genome. 2001
38. Venter JC. Initial sequencing and analysis of the human genome. *Nature.* 2001; 409:860–921. [PubMed: 11237011]
39. Parsons DW, Jones S, Zhang X, Lin JCH, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu I. An Integrated Genomic Analysis of Human Glioblastoma Multiforme. *Science.* 2008; 321(5897):1807. [PubMed: 18772396]
40. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. The consensus coding sequences of human breast and colorectal cancers. *Science.* 2006; 314(5797):268–74. [PubMed: 16959974]
41. Nevins JR, Potti A. Mining gene expression profiles: expression signatures as cancer phenotypes. *Nature reviews. Genetics.* 2007; 8(8):601. [PubMed: 17607306]
42. Coe BP, Chari R, Lockwood WW, Lam WL. Evolving strategies for global gene expression analysis of cancer. *Journal of cellular physiology.* 2008; 217(3):590. [PubMed: 18680120]
43. Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet.* 2006; 7(1):55–65. [PubMed: 16369572]
44. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends in Genetics.* 2008; 24(3):133–141. [PubMed: 18262675]
45. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol.* 2008; 26(10):1135–1145. [PubMed: 18846087]
46. Gingras AC, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. *Nature reviews. Molecular cell biology.* 2007; 8(8):645. [PubMed: 17593931]
47. Cox J, Mann M. Is proteomics the new genomics? *Cell.* 2007; 130(3):395. [PubMed: 17693247]
48. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry.* 2007; 389(4):1017. [PubMed: 17668192]
49. Vienna A, GmbH EB, Linz KHE, Linz A, Innsbruck A. Transforming omics data into context: Bioinformatics on genomics and proteomics raw data. *Electrophoresis.* 2006; 27:2659–2675. [PubMed: 16739231]
50. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002; 30(1):207. [PubMed: 11752295]
51. Sherlock G, Hernandez-Boussard T, Kasarskis A, Binkley G, Matese JC, Dwight SS, Kaloper M, Weng S, Jin H, Ball CA. The Stanford Microarray Database. *Nucleic Acids Res.* 2001; 29(1):152–155. [PubMed: 11125075]
52. Perier RC, Junier T, Bucher P. The Eukaryotic Promoter Database EPD. *Nucleic Acids Res.* 1998; 26(1):353–357. [PubMed: 9399872]
53. Zhao F, Xuan Z, Liu L, Zhang MQ. TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies. *Nucleic Acids Res.* 2005; 33:D103. [PubMed: 15608156]
54. Martens L, Hermjakob H, Jones P, Adamski M, Taylor C. PRIDE: The proteomics identifications database. *Proteomics.* 2005; 5(13):3537–3545. [PubMed: 16041671]

55. Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R, Apweiler R. The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res.* 2004; 32(90001):W313–W317. [PubMed: 15215401]
56. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 2004; 32:D115. [PubMed: 14681372]
57. Camon E, Magrane M, Barrell D, Binns D, Fleischmann W, Kersey P, Mulder N, Oinn T, Maslen J, Cox A. The Gene Ontology Annotation (GOA) Project: Implementation of GO in SWISS-PROT, TrEMBL, and InterPro. *Genome Res.* 2003; 13(4):662–672. [PubMed: 12654719]
58. Flicek P, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T. Ensembl 2008. *Nucleic Acids Res.* 2008; 36(Database issue):D707. [PubMed: 18000006]
59. Sterk P, Kersey PJ, Apweiler R. Genome reviews: standardizing content and representation of information about complete genomes. *OMICS: A Journal of Integrative Biology.* 2006; 10(2):114–118. [PubMed: 16901215]
60. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28(1):27. [PubMed: 10592173]
61. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium *Nat Genet.* 2000; 25(1):25–9.
62. Krieger CJ, Zhang P, Mueller LA, Wang A, Paley S, Arnaud M, Pick J, Rhee SY, Karp PD. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 2004; 32(Database Issue):D438. [PubMed: 14681452]
63. Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahren D, Tsoka S, Darzentas N, Kunin V, Lopez-Bigas N. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res.* 2005; 33(19):6083. [PubMed: 16246909]
64. DeJongh M, Formsma K, Boillot P, Gould J, Rycenga M, Best A. Toward the automated generation of genome-scale metabolic networks in the SEED. *BMC Bioinformatics.* 2007; 8(1):139. [PubMed: 17462086]
65. Ren Q, Chen K, Paulsen IT. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.* 2006
66. Le Novère N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* 2006; 34:D689–D691. [PubMed: 16381960]
67. Sivakumaran S, Hariharaputran S, Mishra J, Bhalla US. The Database of Quantitative Cellular Signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics.* 2003; 19(3):408–415. [PubMed: 12584128]
68. Ji ZL, Chen X, Zhen CJ, Yao LX, Han LY, Yeo WK, Chung PC, Puy HS, Tay YT, Muhammad A. KDBI: Kinetic Data of Bio-molecular Interactions database. *Nucleic Acids Res.* 2003; 31(1):255. [PubMed: 12519995]
69. Platzer A, Perco P, Lukas A, Mayer B. Characterization of protein-interaction networks in tumors. *BMC Bioinformatics.* 2007; 8:224. [PubMed: 17597514]
70. Ramani AK, Bunescu RC, Mooney RJ, Marcotte EM. Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. *Genome Biol.* 2005; 6(5):R40. [PubMed: 15892868]
71. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* 2002; 30(1):303. [PubMed: 11752321]
72. Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G. MINT: a Molecular INteraction database. *FEBS Lett.* 2002; 513(1):135–140. [PubMed: 11911893]
73. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stumpflen V, Mewes HW. The MIPS mammalian protein-protein interaction database. *Bioinformatics.* 2005; 21(6):832–834. [PubMed: 15531608]

74. Lloyd CM, Halstead MDB, Nielsen PF. CellML: its future, present and past. *Prog Biophys Mol Biol.* 2004; 85(2–3):433–450. [PubMed: 15142756]
75. Le Novère N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, Crampin EJ, Halstead M, Klipp E, Mendes P. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol.* 2005; 23:1509–1515. [PubMed: 16333295]
76. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics.* 2003; 19(4):524–531. [PubMed: 12611808]
77. Stromback L, Lambrix P. Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. *Bioinformatics.* 2005; 21(24):4401–4407. [PubMed: 16234320]
78. Shannon PT, Reiss DJ, Bonneau R, Baliga NS. The Gaggle: An open-source software system for integrating bioinformatics software and data sources. *BMC Bioinformatics.* 2006; 7:176. [PubMed: 16569235]
79. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 2003; 13(11):2498–2504. [PubMed: 14597658]
80. Cusick ME, Yu H, Smolyar A, Venkatesan K, Carvunis AR, Simonis N, Rual JF, Borick H, Braun P, Dreze M. Literature-curated protein interaction datasets. 2008
81. Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, Ball CA, Binz PA, Bogue M, Booth T, Brazma A, Brinkman RR, Michael Clark A, Deutsch EW, Fiehn O, Fostel J, Ghazal P, Gibson F, Gray T, Grimes G, Hancock JM, Hardy NW, Hermjakob H, Julian RK Jr, Kane M, Kettner C, Kinsinger C, Kolker E, Kuiper M, Le Novere N, Leebens-Mack J, Lewis SE, Lord P, Mallon AM, Marthandan N, Masuya H, McNally R, Mehrle A, Morrison N, Orchard S, Quackenbush J, Reecy JM, Robertson DG, Rocca-Serra P, Rodriguez H, Rosenfelder H, Santoyo-Lopez J, Scheuermann RH, Schober D, Smith B, Snape J, Stoeckert CJ Jr, Tipton K, Sterk P, Untergasser A, Vandesompele J, Wiemann S. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat Biotechnol.* 2008; 26(8):889–96. [PubMed: 18688244]
82. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet.* 2001; 29(4):365–71. [PubMed: 11726920]
83. Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK Jr, Jones AR, Zhu W, Apweiler R, Aebersold R, Deutsch EW, Dunn MJ, Heck AJ, Leitner A, Macht M, Mann M, Martens L, Neubert TA, Patterson SD, Ping P, Seymour SL, Souda P, Tsugita A, Vandekerckhove J, Vondriska TM, Whitelegge JP, Wilkins MR, Xenarios I, Yates JR 3rd, Hermjakob H. The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol.* 2007; 25(8):887–93. [PubMed: 17687369]
84. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI. An empirical framework for binary interactome mapping. *Nature Methods.* 2008
85. Braun P, Tasan M, Dreze M, Barrios-Rodiles M, Lemmens I, Yu H, Sahalie JM, Murray RR, Roncari L, de Smet AS. An experimentally derived confidence score for binary protein-protein interactions. *Nature methods.* 2009; 6(1):91. [PubMed: 19060903]
86. Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M, Giardine B, Harte RA, Hinrichs AS, Hsu F. The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.* 2008; 36(Database issue):D773. [PubMed: 18086701]
87. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* 2006
88. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. *Nucleic Acids Res.* 2002; 30(1):42. [PubMed: 11752249]
89. Quackenbush J. Microarray Analysis and Tumor Classification. *N Engl J Med.* 2006; 354(23):2463. [PubMed: 16760446]

90. Davis CA, Gerick F, Hintermair V, Friedel CC, Fundel K, Kuffner R, Zimmer R. Reliable gene signatures for microarray classification: assessment of stability and performance. *Bioinformatics*. 2006; 22(19):2356. [PubMed: 16882647]
91. de Souto MCP, Costa IG, de Araujo DSA, Ludermitr TB, Schliep A. Clustering cancer gene expression data: a comparative study. *BMC Bioinformatics*. 2008; 9(1):497. [PubMed: 19038021]
92. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000; 403:503–511. [PubMed: 10676951]
93. Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, Dohner H, Pollack JR. Use of Gene-Expression Profiling to Identify Prognostic Subclasses in Adult Acute Myeloid Leukemia. *N Engl J Med*. 2004; 350(16):1605–1616. [PubMed: 15084693]
94. Naderi A, Teschendorff AE, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG, Robertson JFR, Aparicio S, Ellis IO, Brenton JD. A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene*. 2006; 26(10):1507. [PubMed: 16936776]
95. Brown MPS, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M, Haussler D. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National Academy of Sciences*. 2000; 97(1):262–267.
96. Khan J, Wei JS, Ringnér M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med*. 2001; 7:673–679. [PubMed: 11385503]
97. Herrero J, Valencia A, Dopazo J. A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics*. 2001; 17(2):126–136. [PubMed: 11238068]
98. Geman D. Classifying Gene Expression Profiles from Pairwise mRNA Comparisons. *Stat Appl Genet Mol Biol*. 2004;3.
99. Tan AC, Naiman DQ, Xu L, Winslow RL, Geman D. Simple decision rules for classifying human cancers from gene expression profiles. *Bioinformatics*. 2005; 21(20):3896–3904. [PubMed: 16105897]
100. Xu L, Geman D, Winslow RL. Large-scale integration of cancer microarray data identifies a robust common cancer signature. *BMC Bioinformatics*. 2007; 8:275. [PubMed: 17663766]
101. Price ND, Trent J, El-Naggar AK, Cogdell D, Taylor E, Hunt KK, Pollock RE, Hood L, Shmulevich I, Zhang W. Highly accurate two-gene classifier for differentiating gastrointestinal stromal tumors and leiomyosarcomas. *Proceedings of the National Academy of Sciences*. 2007; 104(9):3414.
102. Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, Cheng CL, Wang CH, Terng HJ, Kao SF. A Five-Gene Signature and Clinical Outcome in Non-Small-Cell Lung Cancer. *N Engl J Med*. 2007; 356(1):11. [PubMed: 17202451]
103. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet*. 2005; 365(9458):488–492. [PubMed: 15705458]
104. Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, Behm FG, Raimondi SC, Relling MV, Patel A. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*. 2002; 1(2):133–143. [PubMed: 12086872]
105. Foekens JA, Atkins D, Zhang Y, Sweep FC, Harbeck N, Paradiso A, Cufer T, Sieuwerts AM, Talantov D, Span PN. Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. *J Clin Oncol*. 2006; 24(11):1665–71. [PubMed: 16505412]
106. Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, Hess K, Lecoche M, Metivier J, Booser D, Ibrahim N. Gene Expression Profiles Predict Complete Pathologic Response to Neoadjuvant Paclitaxel and Fluorouracil Doxorubicin, and Cyclophosphamide Chemotherapy in Breast Cancer. *J Clin Oncol*. 2004; 22(12):2284. [PubMed: 15136595]
107. Dressman HK, Hans C, Bild A, Olson JA, Rosen E, Marcom PK, Liotcheva VB, Jones EL, Vujaskovic Z, Marks J. Gene Expression Profiles of Multiple Breast Cancer Phenotypes and

- Response to Neoadjuvant Chemotherapy. *Clin Cancer Res.* 2006; 12(3):819–826. [PubMed: 16467094]
108. Thuerigen O, Schneeweiss A, Toedt G, Warnat P, Hahn M, Kramer H, Brors B, Rudlowski C, Benner A, Schuetz F. Gene Expression Signature Predicting Pathologic Complete Response With Gemcitabine Epirubicin, and Docetaxel in Primary Breast Cancer. *J Clin Oncol.* 2006; 24(12):1839. [PubMed: 16622258]
 109. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics.* 2007; 23(23):3251. [PubMed: 17644558]
 110. Kim SY, Volsky DJ. PAGE: Parametric Analysis of Gene Set Enrichment. *BMC Bioinformatics.* 2005; 6:144. [PubMed: 15941488]
 111. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences.* 2005; 102(43):15545–15550.
 112. Rhodes DR, Chinnaiyan AM. Integrative analysis of the cancer transcriptome. *Nat Genet.* 2005; 37(6 Suppl):S31–S37. [PubMed: 15920528]
 113. Shen R, Chinnaiyan AM, Ghosh D. Pathway analysis reveals functional convergence of gene expression profiles in breast cancer. *BMC Med Genomics.* 2008; 1:28. [PubMed: 18588682]
 114. Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, Kalyana-Sundaram S, Wei JT, Rubin MA, Pienta KJ. Integrative molecular concept modeling of prostate cancer progression. *Nat Genet.* 2006; 39:41–51. [PubMed: 17173048]
 115. Lee E, Chuang HY, Kim JW, Ideker T, Lee D. Inferring Pathway Activity toward Precise Disease Classification. *PLoS Comput Biol.* 2008; 4(11)
 116. Yu JX, Sieuwerts AM, Zhang Y, Martens JWM, Smid M, Klijn JGM, Wang Y, Foekens JA. Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer. *BMC Cancer.* 2007; 7:182. [PubMed: 17894856]
 117. Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JGM, Foekens JA, Martens JWM. Subtypes of Breast Cancer Show Preferential Site of Relapse. *Cancer Res.* 2008; 68(9):3108. [PubMed: 18451135]
 118. Saal LH, Johansson P, Holm K, Gruvberger-Saal SK, She QB, Maurer M, Koujak S, Ferrando AA, Malmstrom P, Memeo L. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proceedings of the National Academy of Sciences.* 2007; 104(18):7564.
 119. Price, ND.; Edelman, LB.; Lee, I.; Yoo, H.; Hwang, D.; Carlson, G.; Galas, DJ.; Heath, JR.; Hood, L. Systems biology and systems medicine. In: Ginsburg, Geoffrey S.; Willard, Huntington F., editors. *Genomic and Personalized Medicine.* Elsevier; Amsterdam: 2009. p. 74-86.
 120. Hood L, Heath JR, Phelps ME, Lin B. Systems Biology and New Technologies Enable Predictive and Preventative Medicine. *Science.* 2004; 306(5696):640–643. [PubMed: 15499008]
 121. Kitano H. Systems biology: a brief overview. *Science.* 2002; 295(5560):1662–4. [PubMed: 11872829]
 122. Lage K, Karlberg EO, Størling ZM, Ólason P, Pedersen AG, Rigina O, Hinsby AM, Tümer Z, Pociot F, Tommerup N. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol.* 2007; 25:309–316. [PubMed: 17344885]
 123. Price ND, Shmulevich I. Biochemical and statistical network models for systems biology. *Curr Opin Biotechnol.* 2007; 18(4):365–370. [PubMed: 17681779]
 124. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. *Science.* 2005; 308(5721):523–529. [PubMed: 15845847]
 125. Wang RS, Wang Y, Zhang XS, Chen L. Inferring transcriptional regulatory networks from high-throughput data. *Bioinformatics.* 2007; 23(22):3056. [PubMed: 17890736]
 126. Bansal M, Gatta GD, di Bernardo D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics.* 2006; 22(7):815–822. [PubMed: 16418235]

127. Bonneau R. Learning biological networks: from modules to dynamics. *Nat Chem Biol.* 2008; 4(11):658–664. [PubMed: 18936750]
128. Miller-Jensen K, Janes KA, Brugge JS, Lauffenburger DA. Common effector processing mediates cell-specific responses to stimuli. *Nature.* 2007; 448(7153):604. [PubMed: 17637676]
129. Janes KA, Albeck JG, Gaudet S, Sorger PK, Lauffenburger DA, Yaffe MB. A Systems Model of Signaling Identifies a Molecular Basis Set for Cytokine-Induced Apoptosis. *Science.* 2005; 310(5754):1646–1653. [PubMed: 16339439]
130. Gordus A, Krall JA, Beyer EM, Kaushansky A, Wolf-Yadlin A, Sevecka M, Chang BH, Rush J, MacBeath G. Linear combinations of docking affinities explain quantitative differences in RTK signaling. *Mol Syst Biol.* 2009; 5:235. [PubMed: 19156127]
131. Kumar N, Afeyan R, Kim HD, Lauffenburger DA. Multi-Pathway Model Enables Prediction of Kinase Inhibitor Cross-Talk Effects on Migration of Her2-Overexpressing Mammary Epithelial Cells. *Mol Pharmacol.* 2008
132. Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, Thorsson V. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol.* 2006; 7(5):R36. [PubMed: 16686963]
133. Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsson V, Shannon P, Johnson MH, Bare JC. A Predictive Model for Transcriptional Control of Physiology in a Free Living Cell. *Cell.* 2007; 131(7):1354–1365. [PubMed: 18160043]
134. Bansal M, Belcastro V, Ambesi-Impombato A, di Bernardo D. How to infer gene networks from expression profiles. *Mol Syst Biol.* 2007; 3:78. [PubMed: 17299415]
135. Schafer J, Strimmer K. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics.* 2005; 21(6):754–764. [PubMed: 15479708]
136. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet.* 2005; 37:382–390. [PubMed: 15778709]
137. Hartemink AJ. Reverse engineering gene regulatory networks. *Nat Biotechnol.* 2005; 23:554–555. [PubMed: 15877071]
138. Friedman N. Inferring cellular networks using probabilistic graphical models. *Science.* 2004; 303(5659):799–805. [PubMed: 14764868]
139. Shmulevich I, Dougherty ER, Kim S, Zhang W. Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics.* 2002; 18(2):261–274. [PubMed: 11847074]
140. Hashimoto RF, Kim S, Shmulevich I, Zhang W, Bittner ML, Dougherty ER. Growing genetic regulatory networks from seed genes. *Bioinformatics.* 2004; 20(8):1241–1247. [PubMed: 14871865]
141. Pal R, Datta A, Bittner ML, Dougherty ER. Intervention in context-sensitive probabilistic Boolean networks. *Bioinformatics.* 2005; 21(7):1211–1218. [PubMed: 15531600]
142. Soranzo N, Bianconi G, Altafini C. Comparing association network algorithms for reverse engineering of large-scale gene regulatory networks: synthetic versus real data. *Bioinformatics.* 2007; 23(13):1640. [PubMed: 17485431]
143. Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics.* 2004; 20(18):3594–3603. [PubMed: 15284094]
144. Camacho D, Vera LP, Mendes P, Laubenbacher R. Comparison of reverse-engineering methods using an in silico network. *Ann N Y Acad Sci.* 2007; 1115:73–89. [PubMed: 17925358]
145. Margolin AA, Califano A. Theory and limitations of genetic network inference from microarray data. *Ann N Y Acad Sci.* 2007; 1115:51–72. [PubMed: 17925348]
146. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson B. Reconstruction of biochemical networks in microbial organisms. *Nat Rev Microbiol.* 2009; 7(2):129–143. [PubMed: 19116616]
147. Francke C, Siezen RJ, Teusink B. Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol.* 2005; 13(11):550–558. [PubMed: 16169729]
148. Reed JL, Famili I, Thiele I, Palsson BO. Towards multidimensional genome annotation. *Nat Rev Genet.* 2006; 7(2):130–141. [PubMed: 16418748]

149. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson B. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol.* 2007; 3(121)
150. Feist AM, Scholten JCM, Palsson B, Brockman FJ, Ideker T. Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Mol Syst Biol.* 2006; 2(2006.0004)
151. Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Buthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novere N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasie I, Weichart D, Brent R, Broomhead DS, Westerhoff HV, Kuardar B, Penttila M, Klipp E, Palsson BO, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol.* 2008; 26(10):1155–1160. [PubMed: 18846089]
152. Chavali AK, Whittemore JD, Eddy JA, Williams KT, Papin JA. Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. *Mol Syst Biol.* 2008; 4(177)
153. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences.* 2007; 104(6):1777.
154. Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I. The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol.* 2007; 3(135)
155. Hoffmann A, Levchenko A, Scott ML, Baltimore D. The Ikappa B-NF-kappa B Signaling Module: Temporal Control and Selective Gene Activation. *Science.* 2002; 298(5596):1241–1245. [PubMed: 12424381]
156. Papin JA, Hunter T, Palsson BO, Subramaniam S. Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol.* 2005; 6:99–111. [PubMed: 15654321]
157. Papin JA, Palsson BO. Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. *J Theor Biol.* 2004; 227(2):283–297. [PubMed: 14990392]
158. Gianchandani EP, Papin JA, Price ND, Joyce AR, Palsson BO, Albert R. Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Comput Biol.* 2006; 2(8):e101. [PubMed: 16895435]
159. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. *Nature.* 2004; 429(6987):92–96. [PubMed: 15129285]
160. Price ND, Reed JL, Palsson BO. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol.* 2004; 2(11):886–897. [PubMed: 15494745]
161. Vo TD, Greenberg HJ, Palsson BO. Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *J Biol Chem.* 2004; 279(38):39532–39540. [PubMed: 15205464]
162. Thiele I, Price ND, Vo TD, Palsson BO. Candidate Metabolic Network States in Human Mitochondria: IMPACT OF DIABETES, ISCHEMIA, AND DIET. *J Biol Chem.* 2005; 280(12):11683–11695. [PubMed: 15572364]
163. Price ND, Schellenberger J, Palsson BO. Uniform Sampling of Steady-State Flux Spaces: Means to Design Experiments and to Interpret Enzymopathies. *Biophys J.* 2004; 87(4):2172–2186. [PubMed: 15454420]
164. Papin JA, Palsson BO. The JAK-STAT Signaling Network in the Human B-Cell: An Extreme Signaling Pathway Analysis. *Biophys J.* 2004; 87(1):37–46. [PubMed: 15240442]
165. Lee, DS.; Park, J.; Kay, KA.; Christakis, NA.; Oltvai, ZN.; Barabási, AL. The implications of human metabolic network topology for disease comorbidity. *Proceedings of the National Academy of Sciences;* 2008.
166. Ma H, Goryanin I. Human metabolic network reconstruction and its impact on drug discovery and development. *Drug Discov Today.* 2008; 13(9–10):402–408. [PubMed: 18468557]

167. Shlomi T, Cabili MN, Herrgård MJ, Palsson B, Rupp E. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol.* 2008; 26(9):1003–1010. [PubMed: 18711341]
168. Robey IF, Lien AD, Welsh SJ, Baggett BK, Gillies RJ. Hypoxia-Inducible Factor-1 α and the Glycolytic Phenotype in Tumors. *Neoplasia.* 2005; 7(4):324. [PubMed: 15967109]
169. Esteban MA, Maxwell PH. HIF, a missing link between metabolism and cancer. *Nat Med.* 2005; 11(10):1047–8. [PubMed: 16211034]
170. Kroemer G, Pouyssegur J. Tumor Cell Metabolism: Cancer's Achilles' Heel. *Cancer Cell.* 2008; 13(6):472–82. [PubMed: 18538731]
171. Hsu PP, Sabatini DM. Cancer Cell Metabolism: Warburg and Beyond. *Cell.* 2008; 134(5):703–707. [PubMed: 18775299]
172. Galmarini CM, Popowycz F, Joseph B. Cytotoxic Nucleoside Analogues: Different Strategies to Improve their Clinical Efficacy. *Curr Med Chem.* 2008; 15(11):1072–1082. [PubMed: 18473803]
173. Serkova NJ, Spratlin JL, Eckhardt SG. NMR-based metabolomics: Translational application and treatment of cancer. *Current Opinion in Molecular Therapeutics.* 2007; 9(6):572–585. [PubMed: 18041668]
174. Robak T, Lech-Maranda E, Korycka A, Robak E. Purine Nucleoside Analogues as Immunosuppressive and Antineoplastic Agents: Mechanism of Action and Clinical Activity. *Curr Med Chem.* 2006; 13(26):3165–3189. [PubMed: 17168705]
175. Werner SL, Barken D, Hoffmann A. Stimulus Specificity of Gene Expression Programs Determined by Temporal Control of IKK Activity. *Science.* 2005; 309(5742):1857–1861. [PubMed: 16166517]
176. Segre D, Zucker J, Katz J, Lin X, D'Haeseleer P, Rindone WP, Kharchenko P, Nguyen DH, Wright MA, Church GM. From Annotated Genomes to Metabolic Flux Models and Kinetic Parameter Fitting. *OMICS: A Journal of Integrative Biology.* 2003; 7(3):301–316. [PubMed: 14583118]
177. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol.* 2006; 8(11):1195–203. [PubMed: 17060902]
178. Kholodenko BN. Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol.* 2006; 7(3):165–76. [PubMed: 16482094]
179. Covert MW, Leung TH, Gaston JE, Baltimore D. Achieving Stability of Lipopolysaccharide-Induced NF- κ B Activation. *Science.* 2005; 309(5742):1854–1857. [PubMed: 16166516]
180. Geva-Zatorsky N, Rosenfeld N, Itzkovitz S, Milo R, Sigal A, Dekel E, Yarnitzky T, Liron Y, Polak P, Lahav G, Alon U. Oscillations and variability in the p53 system. *Mol Syst Biol.* 2006; 2:2006–0033. [PubMed: 16773083]
181. Ma L, Wagner J, Rice JJ, Hu W, Levine AJ, Stolovitzky GA. A plausible model for the digital response of p53 to DNA damage. *Proc Natl Acad Sci U S A.* 2005; 102(40):14266–71. [PubMed: 16186499]
182. Hennessy BT, Smith DL, Ram PT, Lu Y, Mills GB. Exploiting the PI3K/AKT pathway for cancer drug discovery. *Nat Rev Drug Discov.* 2005; 4(12):988–1004. [PubMed: 16341064]
183. Asthagiri AR, Lauffenburger DA. A computational study of feedback effects on signal dynamics in a mitogen-activated protein kinase (MAPK) pathway model. *Biotechnol Prog.* 2001; 17(2):227–39. [PubMed: 11312698]
184. Brightman FA, Fell DA. Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells. *FEBS Lett.* 2000; 482(3):169–74. [PubMed: 11024454]
185. Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol.* 2002; 20(4):370–5. [PubMed: 11923843]
186. Christopher R, Dhiman A, Fox J, Gendelman R, Haberitcher T, Kagle D, Spizz G, Khalil IG, Hill C. Data-Driven Computer Simulation of Human Cancer Cell. *Ann N Y Acad Sci.* 2004; 1020(1):132–153. [PubMed: 15208190]
187. Li H, Fan X, Houghton J. Tumor microenvironment: the role of the tumor stroma in cancer. *J Cell Biochem.* 2007; 101(4):805–15. [PubMed: 17226777]

188. Bellomo N, Preziosi L. Modelling and Mathematical Problems Related to Tumor Evolution and Its Interaction with the Immune System. *Mathematical and Computer Modelling*. 2000; 32(3): 413–452.
189. Araujo RP, McElwain DL. A history of the study of solid tumour growth: the contribution of mathematical modelling. *Bull Math Biol*. 2004; 66(5):1039–1091. [PubMed: 15294418]
190. Araujo RP, McElwain DLS. A history of the study of solid tumour growth: The contribution of mathematical modelling. *Bull Math Biol*. 2004; 66(5):1039–1091. [PubMed: 15294418]
191. Quaranta V, Rejniak KA, Gerlee P, Anderson AR. Invasion emerges from cancer cell adaptation to competitive microenvironments: Quantitative predictions from multiscale mathematical models. *Semin Cancer Biol*. 2008
192. Alarcón T, Byrne HM, Maini PK. A mathematical model of the effects of hypoxia on the cell-cycle of normal and cancer cells. *J Theor Biol*. 2004; 229(3):395–411. [PubMed: 15234206]
193. Ambrosi D, Mollica F. The role of stress in the growth of a multicell spheroid. *J Math Biol*. 2004; 48(5):477–499. [PubMed: 15133619]
194. Anderson ARA. A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion. *Math Med Biol*. 2005; 22(2):163–186. [PubMed: 15781426]
195. Johnston MD, Edwards CM, Bodmer WF, Maini PK, Chapman SJ. Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proceedings of the National Academy of Sciences*. 2007; 104(10):4008.
196. d'Onofrio A, Tomlinson IPM. A nonlinear mathematical model of cell turnover, differentiation and tumorigenesis in the intestinal crypt. *J Theor Biol*. 2007; 244(3):367–374. [PubMed: 17049944]
197. Byrne H, Preziosi L. Modelling solid tumour growth using the theory of mixtures. *Math Med Biol*. 2003; 20(4):341–366. [PubMed: 14969384]
198. Breward CJW, Byrne HM, Lewis CE. The role of cell-cell interactions in a two-phase model for avascular tumour growth. *J Math Biol*. 2002; 45(2):125–152. [PubMed: 12181602]
199. Macklin P, Lowengrub J. Nonlinear simulation of the effect of microenvironment on tumor growth. *J Theor Biol*. 2007; 245(4):677–704. [PubMed: 17239903]
200. Alarcon T, Byrne HM, Maini PK. A multiple scale model for tumor growth. *Multiscale Modeling & Simulation*. 2005; 3(2):440–475.
201. Spencer SL, Berryman MJ, García JA, Abbott D. An ordinary differential equation model for the multistep transformation to cancer. *J Theor Biol*. 2004; 231(4):515–524. [PubMed: 15488528]
202. Baum M, Chaplain MAJ, Anderson ARA, Douek M, Vaidya JS. Does breast cancer exist in a state of chaos? *Eur. J Cancer*. 1999; 35(6):886–891.
203. Rew DA. Tumour biology, chaos and non-linear dynamics. *Eur J Surg Oncol*. 1999; 25(1):86–89. [PubMed: 10188862]
204. Deisboeck TS, Yoon J, Costa J. In silico cancer modeling: is it ready for prime time? *Nature Clinical Practice Oncology*. 2008; 6(1):34–42.
205. Mansury Y, Kimura M, Lobo J, Deisboeck TS. Emerging patterns in tumor systems: simulating the dynamics of multicellular clusters with an agent-based spatial agglomeration model. *J Theor Biol*. 2002; 219(3):343–70. [PubMed: 12419662]
206. Little MP, Vineis P, Li G. A stochastic carcinogenesis model incorporating multiple types of genomic instability fitted to colon cancer data. *J Theor Biol*. 2008
207. Dingli D, Traulsen A, Pacheco JM. Stochastic dynamics of hematopoietic tumor stem cells. *Cell Cycle*. 2007; 6(4):461–6. [PubMed: 17329969]
208. Ostby I, Oyehaug L, Steen HB. A stochastic model of cancer initiation including a bystander effect. *J Theor Biol*. 2006; 241(4):751–64. [PubMed: 16499930]
209. Kansal AR, Torquato S, Harsh Iv GR, Chiocca EA, Deisboeck TS. Simulated brain tumor growth dynamics using a three-dimensional cellular automaton. *J Theor Biol*. 2000; 203(4):367–82. [PubMed: 10736214]
210. Mallet DG, De Pillis LG. A cellular automata model of tumor-immune system interactions. *J Theor Biol*. 2006; 239(3):334–350. [PubMed: 16169016]

211. Alarcón T, Byrne HM, Maini PK. A cellular automaton model for tumour growth in inhomogeneous environment. *J Theor Biol.* 2003; 225(2):257–274. [PubMed: 14575659]
212. Bauer AL, Jackson TL, Jiang Y. A Cell-Based Model Exhibiting Branching and Anastomosis during Tumor-Induced Angiogenesis. *Biophys J.* 2007; 92(9):3105. [PubMed: 17277180]
213. Patel AA, Gawlinski ET, Lemieux SK, Gatenby RA. A Cellular Automaton Model of Early Tumor Growth and Invasion: The Effects of Native Tissue Vascularity and Increased Anaerobic Tumor Metabolism. *J Theor Biol.* 2001; 213(3):315–331. [PubMed: 11735284]
214. Mansury Y, Deisboeck TS. The impact of “search precision” in an agent-based tumor model. *J Theor Biol.* 2003; 224(3):325–337. [PubMed: 12941591]
215. McDougall SR, Anderson ARA, Chaplain MAJ. Mathematical modelling of dynamic adaptive tumour-induced angiogenesis: Clinical implications and therapeutic targeting strategies. *J Theor Biol.* 2006; 241(3):564–589. [PubMed: 16487543]
216. Anderson AR. A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion. *Math Med Biol.* 2005; 22(2):163–86. [PubMed: 15781426]
217. Zhang L, Athale CA, Deisboeck TS. Development of a three-dimensional multiscale agent-based tumor model: simulating gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer. *J Theor Biol.* 2007; 244(1):96–107. [PubMed: 16949103]
218. Anderson AR, Weaver AM, Cummings PT, Quaranta V. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. *Cell.* 2006; 127(5):905. [PubMed: 17129778]
219. Ohno R, Nakamura Y. Prediction of response to imatinib by cDNA microarray analysis 2003. *Semin Hematol.* :42–49.

Figure 1.
Molecular and physiological complexity in cancer.

Figure 2.
Biological scales and potential modeling approaches.

Figure 3.
Schematic overview of statistical modeling in human disease.

Figure 4.
Comparison of biochemical reaction network and statistical network models.

Figure 5. Mathematical representation of reaction links in biochemical networks.

Table 1

Databases and software for building systems models of cancer.

Resource	Database	Refs
<i>Genome sequence data</i>	Ensembl	[58]
	UCSC Genome Browser	[86]
<i>Genome annotation data</i>		
Genetic elements	Entrez Gene	[87]
	Gene Ontology Annotation (GOA) Database	[55, 57]
	Universal Protein (UniProt) Knowledgebase	[56]
	Genome Reviews	[59]
Biochemical pathways & functional associations	Kyoto Encyclopedia of Genes and Genomes (KEGG)	[60, 88]
	Gene Ontology (GO)	[61]
	The SEED	[64]
	MetaCyc	[62]
	BioCyc	[63]
	TransportDB	[65]
Regulatory Sequences	Eukaryotic Promoter Database (EPD)	[52]
	Transcriptional Regulatory Element Database (TRED)	[53]
<i>Model, model parameter repositories</i>		
	Kinetic Data of Bio-molecular Interactions Database	[68]
	BioModels Database	[66]
	Database of Quantitative Cellular Signaling (DQCCS)	[67]
<i>Protein interaction networks</i>		
	Database of Interacting Proteins (DIP)	[71]
	Molecular INTERaction Database (MINT)	[72]
	Mammalian Protein-protein Interaction Database (MIPS)	[73]
<i>High-throughput genome-scale data</i>		
Transcriptomics	Gene Expression Omnibus (GEO)	[50]
	Stanford Microarray Database (SMD)	[51]
Proteomics	Proteomics Identifications Database (PRIDE)	[54]
<i>Visualization and data management software packages</i>		
	Cytoscape	[79]
	The Gaggle	[78]

Table 2

Predicted tissue-specific activity of metabolic genes in the human metabolic reconstruction (data from [167]).

Tissue	Active (of 1165)	Inactive	Undetermined
<i>Liver</i>	12.4%	19.2%	68.4%
<i>Brain</i>	7.0%	32.1%	60.9%
<i>Heart</i>	4.2%	22.0%	73.8%
<i>Kidney</i>	12.9%	21.5%	65.7%
<i>Lung</i>	3.9%	33.6%	62.5%
<i>Pancreas</i>	3.8%	35.5%	60.8%
<i>Prostate</i>	2.8%	30.0%	67.1%
<i>Spleen</i>	1.1%	38.5%	60.4%
<i>Thymus</i>	1.0%	37.3%	61.6%
<i>Skeletal Muscle</i>	4.6%	28.0%	67.4%